

11. In a game, a player S stands facing the square with \$0. He draws one of the 3 value cards at random and makes a move according to the value on the card. He replaces the card, draws again and makes the second move. He will be awarded the amount on the final square he lands on. What is the probability that he will win
- (a) \$50?
(b) \$1?
(c) \$0?
- (Hint: Draw a probability tree diagram to represent the outcomes of the game.)

\$0	\$1	\$0	\$1	\$50
-----	-----	-----	-----	------

S

Card value	Action
-1	Move 1 step to the left
0	Do not move
1	Move 1 step to the right

12. In a large population, 25% of the people are smokers. Among the smokers, it was found that 60% were tested positive for a certain type of cancer while only 20% of the non-smokers were tested positive for the disease. A person is chosen at random from the population.
- (a) What is the probability that the person is a non-smoker and is tested negative?
(b) What is the probability that the person is tested positive?

ACTIVITY

ACTIVITY 1

Objective

To explore creating cumulative frequency curves using worksheet software

Procedure

1. Create the following grouped frequency table in columns A and B using worksheet software:

	A	B
1	Height x in cm	Number of plants
2	$15 < x \leq 18$	16
3	$18 < x \leq 21$	20
4	$21 < x \leq 24$	22
5	$24 < x \leq 27$	19
6	$27 < x \leq 30$	13
7	$30 < x \leq 33$	6
8	$33 < x \leq 36$	4

2. Create a new column C as shown below. To generate cumulative frequency in column D using values in column B, define functions in cells D2 to D8 as follows:
- For cell D2, key in '=sum(B\$2:B2)'
 - For cells D3 to D8, copy and paste content of cell D2 into these cells.

	A	B	C	D
1	Height x in cm	Number of plants	$x \leq$	Cumulative frequency
2	$15 < x \leq 18$	16	18	
3	$18 < x \leq 21$	20	21	
4	$21 < x \leq 24$	22	24	
5	$24 < x \leq 27$	19	27	
6	$27 < x \leq 30$	13	30	
7	$30 < x \leq 33$	6	33	
8	$33 < x \leq 36$	4	36	

3. Apply **Chart Wizard** to cells D2 to D8 and choose **Chart Type XY(Scatter)** to generate a cumulative frequency curve.
4. Click on any of the points generated and the dialog box **Format Data Series** will pop up. On the tab **Pattern**, click '**Automatic**' for Line and check the box for '**Smoothed line**'. The cumulative frequency curve will then be generated.

ACTIVITY 2

Objective

To explore how a worksheet

- generates quartiles, medians and percentiles
- picks up the lowest and the highest values from a data set

Description

Worksheet software has many built-in statistical features. In this activity we explore the use of the following functions:

- Median – returns a value for median
- Quartile – returns values for lower and upper quartiles
- Percentile – returns a value for a percentage specified
- Min – returns the lowest value among the data
- Max – returns the highest value among the data

Context

The marks obtained by 40 students in a test are recorded below.

87	72	48	93	67	90	86	64	52	29
12	59	71	52	39	19	77	87	25	44
65	42	82	73	70	78	69	42	52	73
63	50	74	44	50	51	82	73	78	29

Procedure

1. Prepare the following worksheet and enter the above scores into it:

	A	B	C	D	E	F	G	H	I	J
1	87	72	48	93	67	90	86	64	52	29
2	12	59	71	52	39	19	77	87	25	44
3	65	42	82	73	70	78	69	42	52	73
4	63	50	74	44	50	51	82	73	78	29
5										
6	Median =									
7	Lower quartile =									
8	Upper quartile =									
9	35 th percentile =									
10	Lowest score =									
11	Highest score =									

2. To calculate all the values from median to the highest score, follow the instructions as shown below:

To find	Key	Into cells
Median	=median(A1:J4)	D6
1 st quartile	=quartile(A1:J4,1)	D7
3 rd quartile	=quartile(A1:J4,3)	D8
35 th percentile	=percentile(A1:J4,0.35)	D9
Lowest score	=min(A1:J4)	D10
Highest score	=max(A1:J4)	D11

3. Vary the values of some of the scores in cells A1 to J4 and see what happens to all the statistics cells (i.e. D6 to D11).

ACTIVITY 3

Objective

To explore how a worksheet calculates variance and standard deviation

Description

Worksheet software generates variance and standard deviation. The built-in functions provided are

- VARP – returns a value for a population variance
- STDEVP – returns a value for a population standard deviation

Context

We use the same context as in Activity 2 on the marks obtained by 40 students in a test.

87	72	48	93	67	90	86	64	52	29
12	59	71	52	39	19	77	87	25	44
65	42	82	73	70	78	69	42	52	73
63	50	74	44	50	51	82	73	78	29

Procedure

1. Prepare the following worksheet and enter the above scores into it:

	A	B	C	D	E	F	G	H	I	J
1	87	72	48	93	67	90	86	64	52	29
2	12	59	71	52	39	19	77	87	25	44
3	65	42	82	73	70	78	69	42	52	73
4	63	50	74	44	50	51	82	73	78	29
5										
6										
7	Variance =									
8	Standard deviation =									

2. To calculate variance and standard deviation, follow the instructions as shown below:

To find	Key	Into cells
Variance	=VARP(A1:J4)	D7
Standard deviation	=STDEVP(A1:J4)	D8

How different are the 2 sets of statistics?

3. Vary the values of some of the scores in cells A1 to J4 and see what happens to all the statistics cells (i.e. D7 and D8).

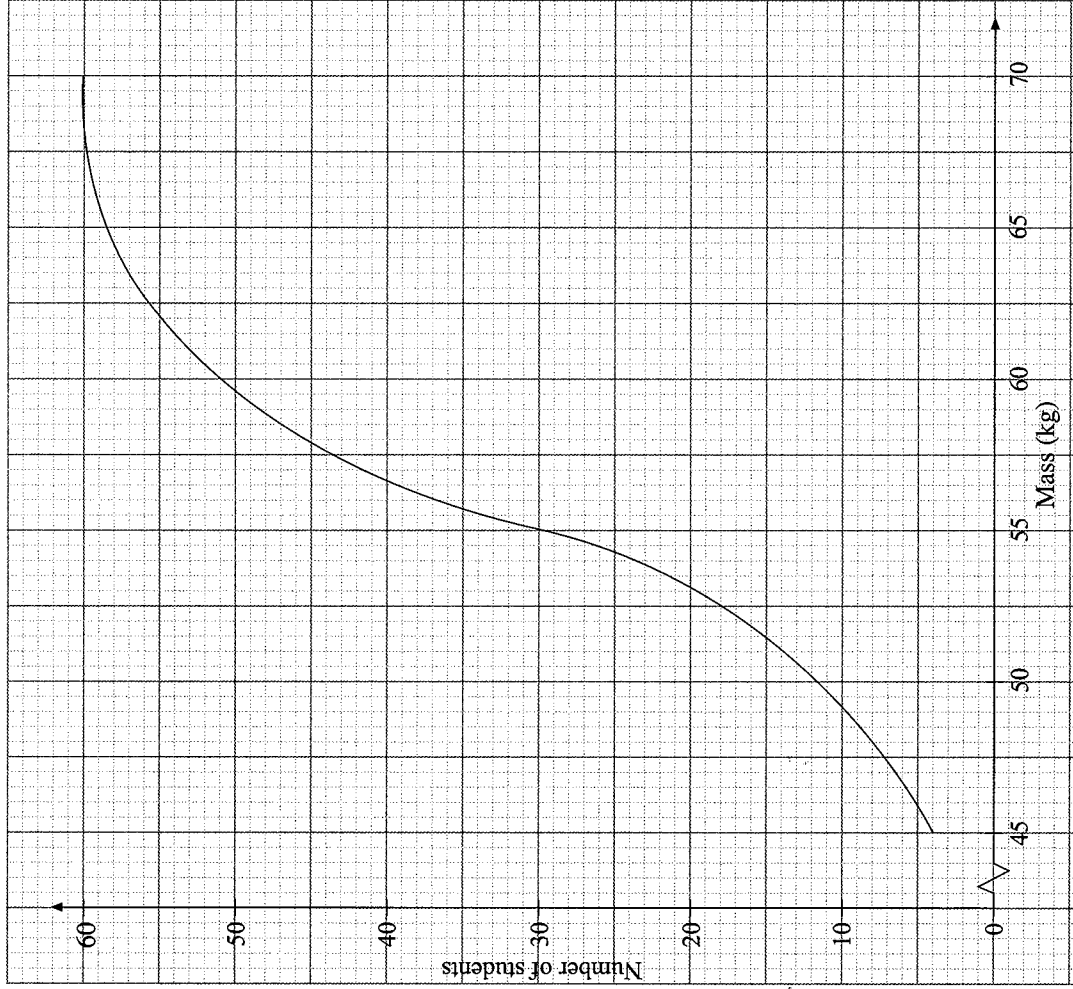
EXERCISE

EXERCISE 1

1. The distances travelled, in kilometres, by 60 salesmen of a company in a day are shown in the cumulative frequency table given.
- (a) How many salesmen travelled more than 25 km a day?
- (b) What percentage of the salesmen travelled 20 km or less?
- (c) How many salesmen travelled between 25 km and 30 km?

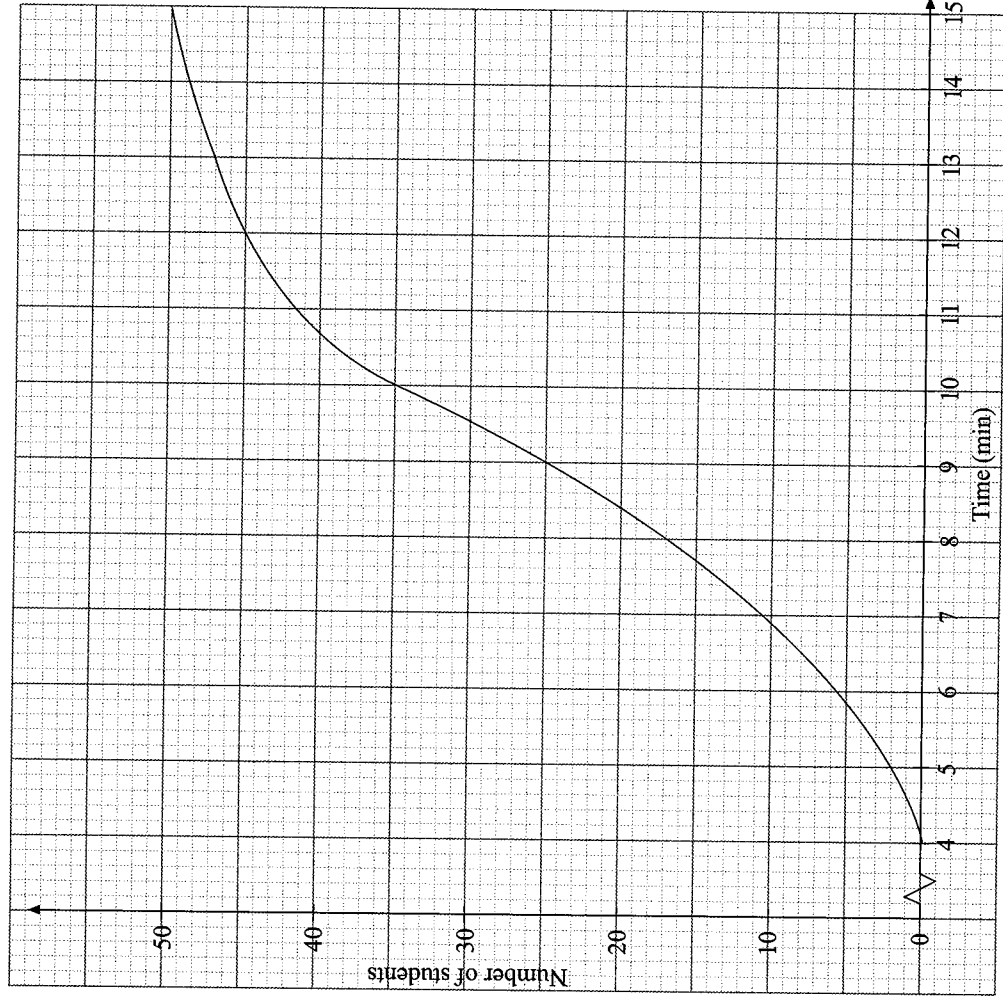
Distance (in km)	Cumulative Frequency
10 or less	10
15 or less	22
20 or less	33
25 or less	41
30 or less	53
35 or less	60

2. The masses of 60 students are shown in the cumulative frequency curve below.



- (a) What percentage of the students weigh more than 45 kg but at most 55 kg?
- (b) How many students weigh more than 70 kg?
- (c) How many students are 48 kg or less?
- (d) If students who are more than 62 kg need to go for the 'TAF' programme, how many students fall into this category?

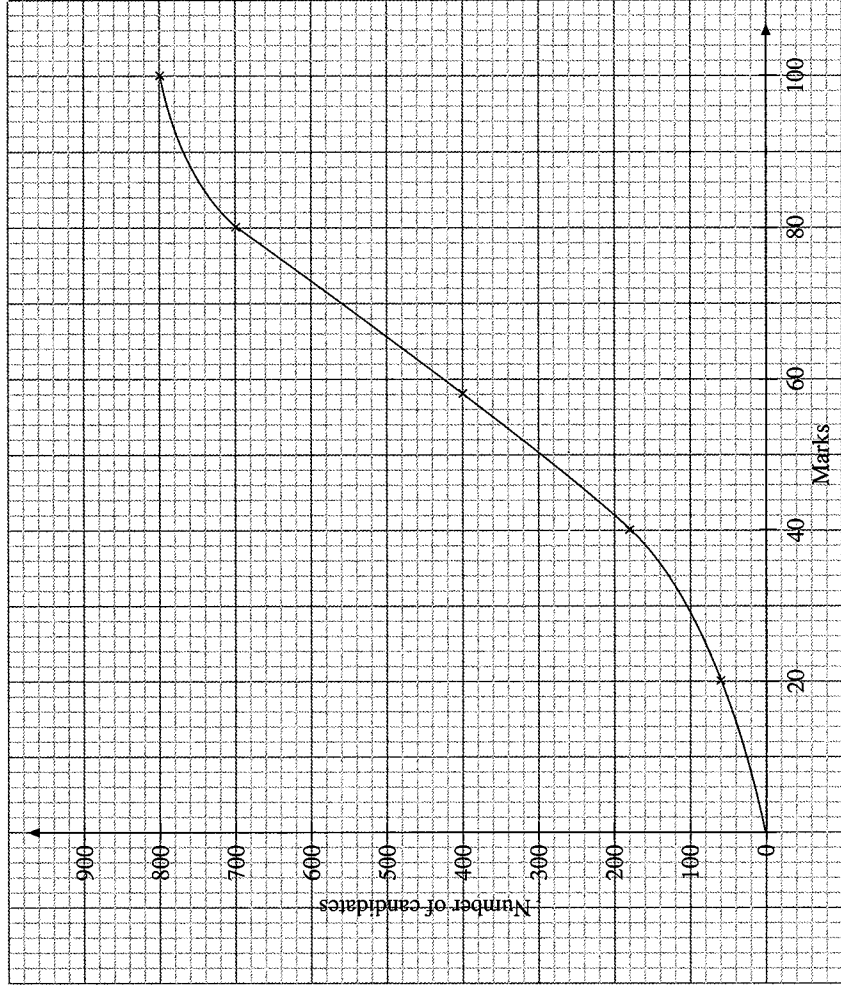
3. The cumulative frequency curve shows the time (in minutes) taken by 50 students to complete a puzzle.



- (a) How many students completed the puzzle within 10 min?
- (b) How many students took between 8 min and 12 min to complete the puzzle?
- (c) What percentage of the students took more than 13 min to complete the puzzle?

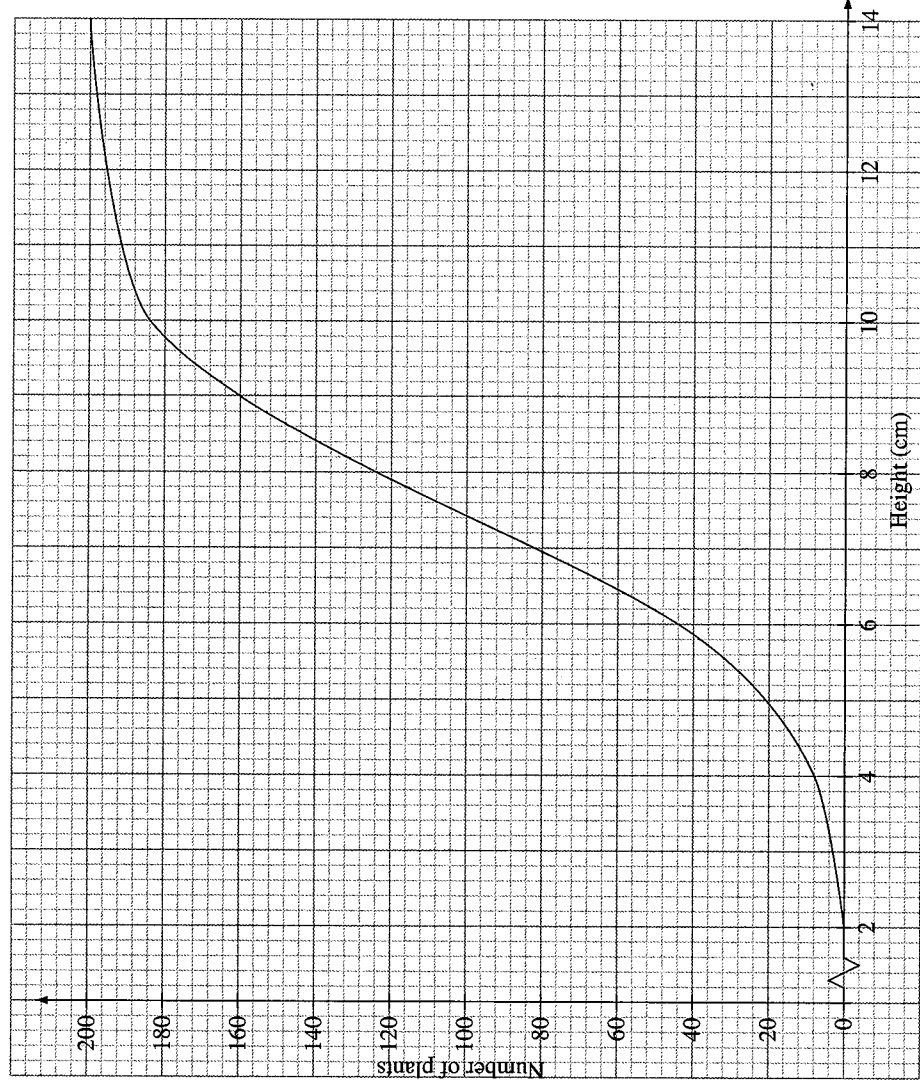
EXERCISE 2

1. The marks scored by 800 pupils in an examination are shown in the cumulative frequency curve below.



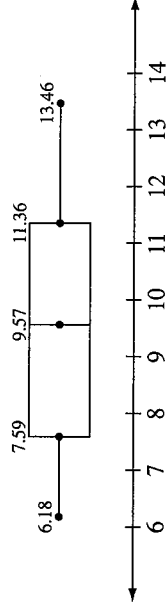
- Using the curve, find
- the median mark,
 - the lower quartile,
 - the upper quartile,
 - the interquartile range,
 - the 80th percentile.

2. The heights of 200 plants of the same species are represented by the cumulative frequency curve below.



- Use the graph to estimate
- the median mark,
 - the lower quartile,
 - the upper quartile,
 - the interquartile range,
 - the 60th percentile.

3. 14 vehicles were checked for carbon monoxide emission (in g/m) and the findings were summarised in the box-and-whisker plot below.



From the box-and-whisker plot, state

- the maximum value,
 - the median,
 - the 1st quartile,
 - the 3rd quartile.
- Hence determine the interquartile range.

EXERCISE

EXERCISE 3

1. Find the variance and standard deviation for each of the following data sets, giving your answers correct to 2 decimal places:

(a) 10, 6, 12, 13, 20, 28, 19, 13, 11, 21, 23

(b) 2.5, 3.14, 6.43, 9.98, 3.33, 4.25, 5, 7.08, 9.5, 6.4

2. The scores of a short quiz taken by 100 students were kept in the frequency table below.

Scores	0	1	2	3	4	5	6	7	8	9	10
Frequency	5	2	4	9	11	18	24	8	11	5	3

Calculate
(a) the mean score,
(b) the variance,
(c) the standard deviation.

3. The number of articles collected by the students was tabulated in the grouped frequency table given. Estimate the standard deviation.

Number of articles (x)	Number of students (f)
$40 \leq x < 60$	7
$60 \leq x < 80$	12
$80 \leq x < 100$	16
$100 \leq x < 120$	8
$120 \leq x < 140$	7

4. Two factories A and B are manufacturing the same electronic components which are subject to quality checks. Rejects are the components that fail to meet the quality specifications. The number of rejects for the past 12 months is tabulated below.

Month	Number of rejects from Factory A	Number of rejects from Factory B
January	50	45
February	64	80
March	45	77
April	50	100
May	65	60
June	86	95
July	45	45
August	78	88
September	46	87
October	50	59
November	43	88
December	62	76

(a) Calculate the monthly average of rejects of each of the 2 factories.

(b) Calculate the standard deviation of rejects of each of the 2 factories. Give your answers correct to 2 decimal places.

(c) Based on their means and standard deviations, what conclusion can be drawn on the quality of the electronic components?

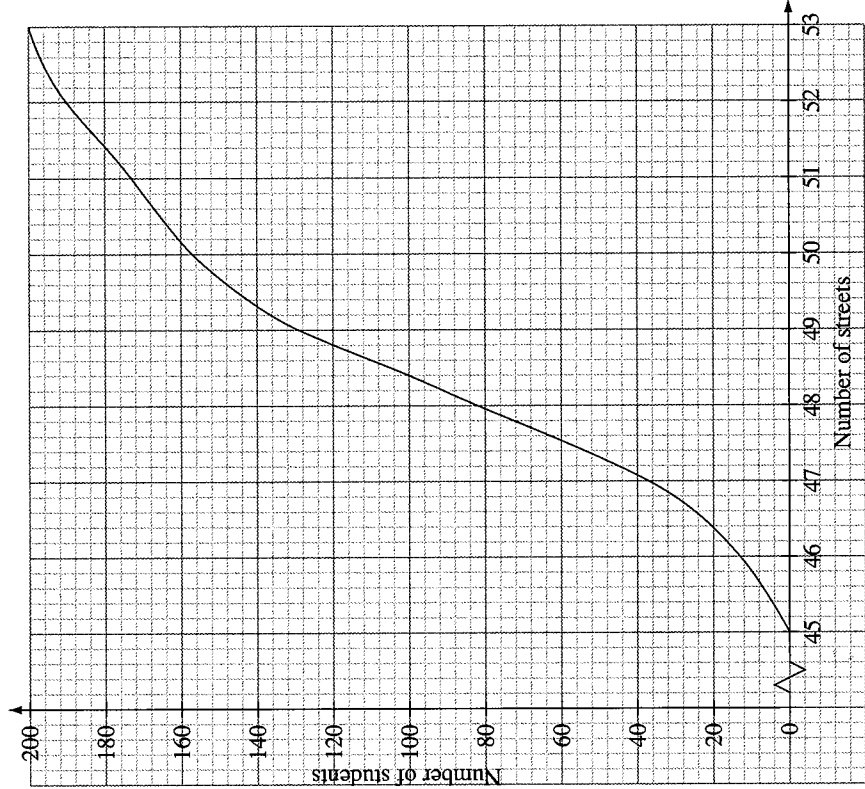
REVISION EXERCISE 6

1. In a contest, 200 students had to mark correctly as many names of the streets as they could on a map of Singapore. The results are given in the following table:

No. of streets correctly marked (x)	46	47	48	49	50	51	52	53
No. of students	13	24	45	47	28	16	17	10

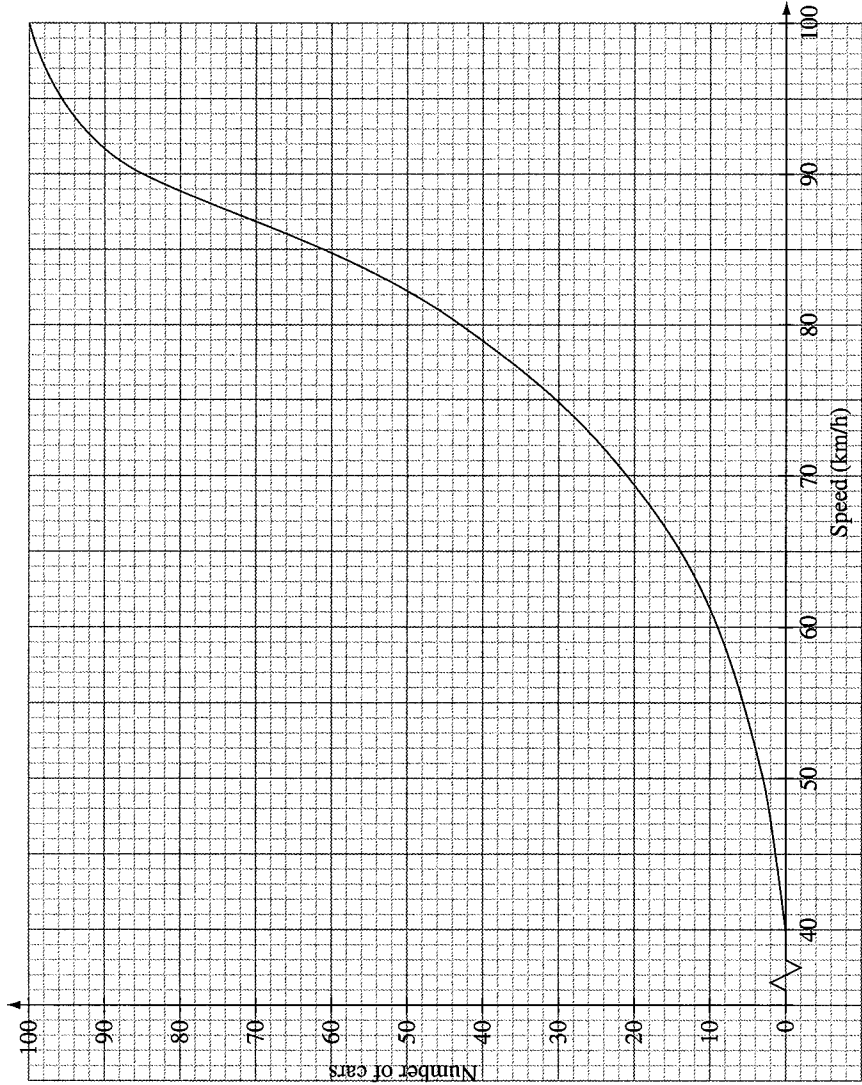
- (a) Find the mean.
- (b) Find the standard deviation.

The results of the students in the contest are represented in the following cumulative frequency curve:



- (c) Find the median.
- (d) Find the percentage of students who marked 48 and less street names correctly.

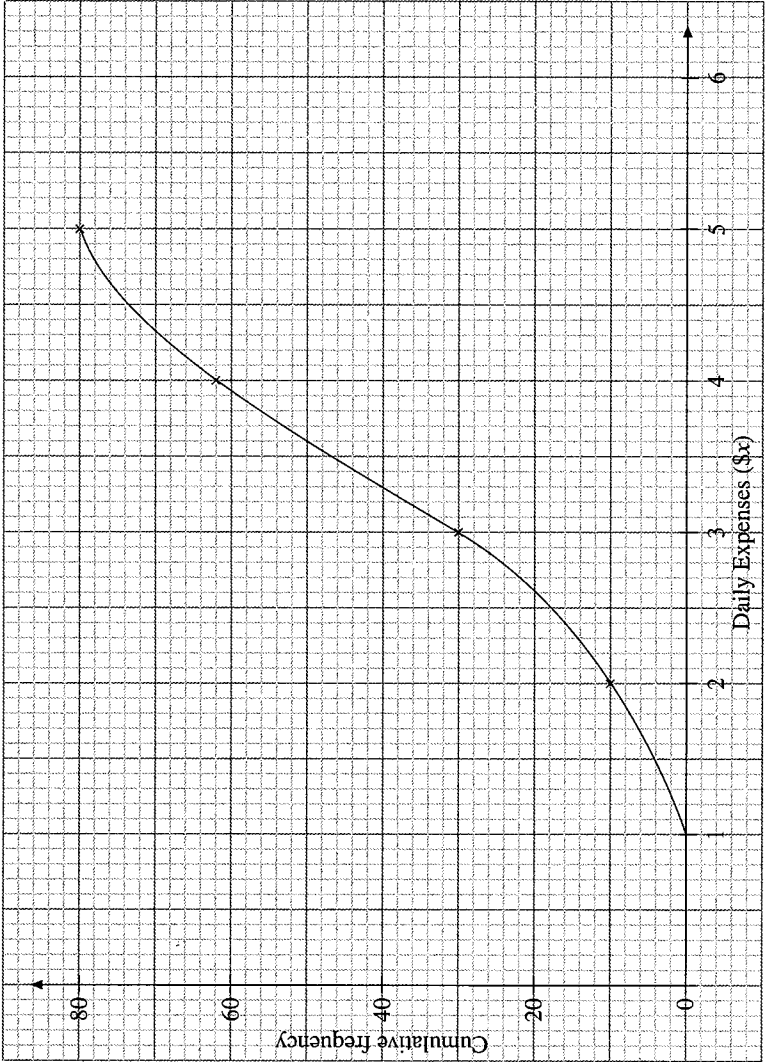
2. The cumulative frequency curve shows the distribution of the speeds, in km/h, of 100 cars.



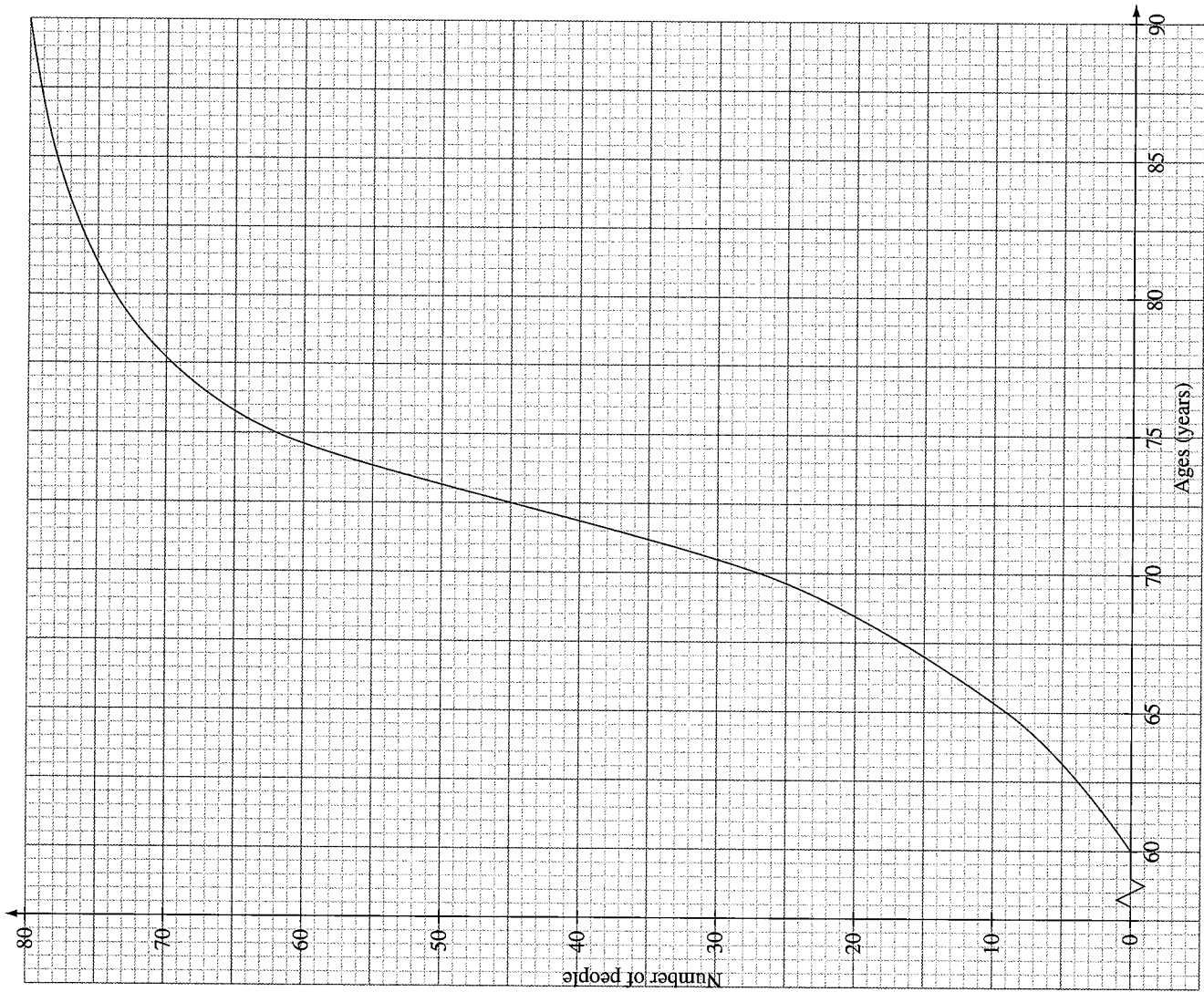
- (a) Use the graph to estimate
- (i) the median speed,
 - (ii) the interquartile range,
 - (iii) the number of cars with a speed not more than 67 km/h.
- (b) A car travelling at a speed exceeding 85 km/h can be issued a traffic summons for speeding. Find the number of cars that are likely to be issued a traffic summons.

3. The graph below is a cumulative frequency curve showing the daily expenses of 80 students. Use the graph to find, as accurately as possible,
- (a) the median,
 - (b) the interquartile range,
 - (c) the number of students who spend \$2.20 or more,
 - (d) the number of students who spend at least \$3.00 but less than \$4.50,
 - (e) the values of a , b and c in the table below.

Daily Expenses (\$x)	$x < 1$	$1 \leq x < 2$	$2 \leq x < 3$	$3 \leq x < 4$	$4 \leq x < 5$
No. of students	0	10	a	b	c



4. The cumulative frequency curve shows the distribution of ages of 80 elderly people in a nursing home.



Use the cumulative frequency curve to estimate

- (a) the median age of the group,
- (b) the upper quartile,
- (c) the 60th percentile,
- (d) the number of elderly people who are at least 82 years old.

5. The masses of 100 men are recorded in the table given.
(a) Copy and complete the table.

Mass (in kg)	Frequency (<i>f</i>)	Mid-value (<i>x</i>)	<i>fx</i>	<i>x</i> ²	<i>fx</i> ²
53.5 – 57.5	3				
57.5 – 61.5	10				
61.5 – 65.5	22				
65.5 – 69.5	34				
69.5 – 73.5	20				
73.5 – 77.5	11				
Total	100				

- (b) Calculate the mean mass.

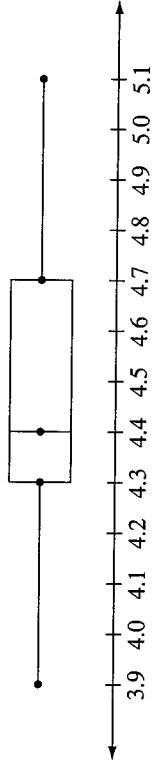
- (c) Calculate the standard deviation.

6. The cumulative distribution of marks gained by a group of 200 pupils in an examination is given in the table below.

Marks	10	20	30	40	50	60	70	80	90	100
No. of pupils scoring this mark or less	12	23	38	59	83	116	155	179	194	200

- (a) How many students scored more than 40 but 70 or less?
(b) If the pass mark is 50, what percentage of the students will pass the examination?

7. The box-and-whisker plot given below shows the summary of masses of a group of babies 3 months old.



- Calculate
(a) the median mass,
(b) the 1st quartile,
(c) the 3rd quartile,
(d) the minimum mass.
What is the interquartile range?

9. The table below shows the salaries of 100 workers in a factory.

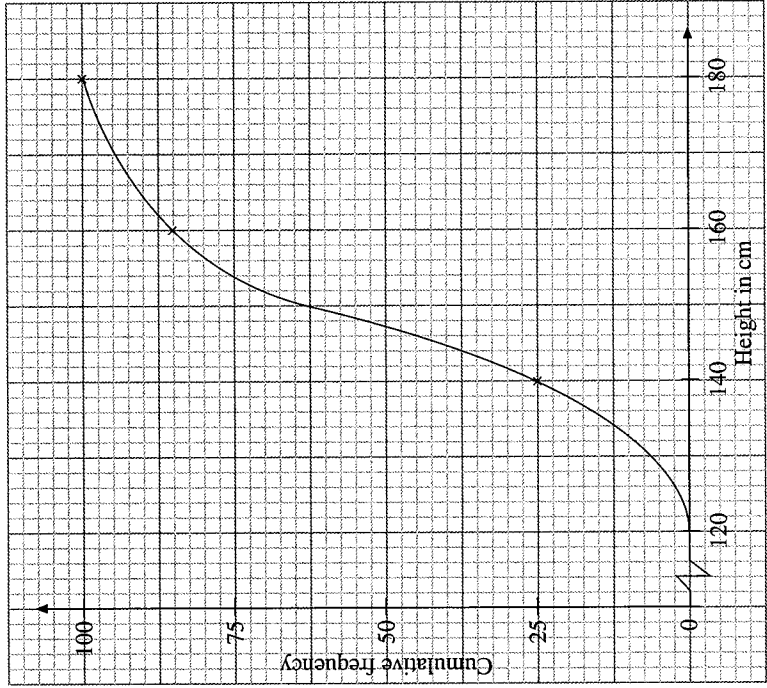
Salary (in thousands)	No. of people
$0 < x \leq 0.5$	5
$0.5 < x \leq 1.0$	9
$1.0 < x \leq 1.5$	18
$1.5 < x \leq 2.0$	24
$2.0 < x \leq 2.5$	27
$2.5 < x \leq 3.0$	13
$3.0 < x \leq 3.5$	4

Calculate the variance and standard deviation of the salaries.

10. The table shows the distribution of examination marks obtained by 200 candidates.
- Calculate
- (a) the mean mark,
 - (b) the variance,
 - (c) the standard deviation.

Marks	Frequency
1 – 10	2
11 – 20	7
21 – 30	15
31 – 40	21
41 – 50	30
51 – 60	41
61 – 70	39
71 – 80	24
81 – 90	15
91 – 100	6

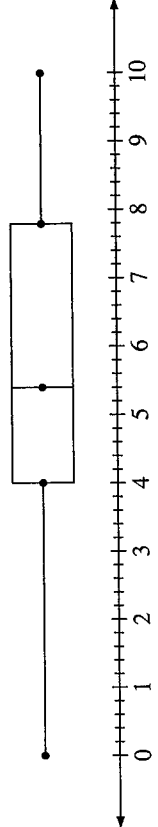
8. The cumulative frequency curve represents the distribution of heights of 100 children.



Using the curve, estimate

- (a) the median height,
- (b) the lower quartile,
- (c) the upper quartile,
- (d) the lowest value,
- (e) the highest value.

11. The box-and-whisker plot below is the summary of scores obtained by contestants in a game.



- (a) Estimate
- (i) the maximum value,
 - (ii) the minimum value,
 - (iii) the lower quartile,
 - (iv) the upper quartile,
 - (v) the median.
- (b) If there were 160 contestants, how many of them would score 4 points and above?

12. Henry and Carl go jogging every morning. The distances they jog have the following statistics:

	Mean	Standard deviation
Henry's distance	5.25 km	2 km
Carl's distance	4.20 km	0.75 km

- (a) What would you say about the distances they jog based only on their mean distance?
- (b) From the means and the standard deviations in the above table, what can you conclude about the distances they jog?

13. 2 classes of 20 students each sat for an examination and obtained the following marks:

Class A				
84	63	90	68	42
43	31	60	88	70
25	40	32	37	79
66	55	65	35	42

Class B				
63	66	62	66	80
55	72	77	66	58
66	68	44	60	70
66	76	75	71	74

- (a) Find the range of marks for class A and class B respectively.
- (b) Calculate the mean and the standard deviation for each class.
- (c) Based on the means and standard deviations of these 2 classes, how would you compare the 2 classes?